CED WORKING PAPER

AN AUTOMATED WEATHER/YIELD SYSTEM: DEVELOPMENT, STRUCTURE, AND FUNCTION

By

Michael D. Weiss *

Forecast Support Group
April 1979

COMMODITY
ECONOMICS DIVISION

ECONOMICS, STATISTICS, and COOPERATIVES SERVICE
U.S. DEPARTMENT OF AGRICULTURE

AN AUTOMATED WEATHER/YIELD SYSTEM:

DEVELOPMENT, STRUCTURE, AND FUNCTION

By

Michael D. Weiss *

Forecast Support Group
April 1979

* The author wishes to express his appreciation to Dr. Wayne Boutwell and Dr. Abner Womack for their support and encouragement during the development of the system described here.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

The winter of 1976-77 brought with it not only extreme cold, but also the threat of drought in significant portions of U.S. crop-producing regions. With this backdrop, Agriculture Secretary Bergland, new in office, announced that USDA would intensify its efforts to quantify the effects of weather on agriculture.

In February 1977, as part of his Agency's response to this need, the author was charged with the task of developing improved operational forecasting models reflecting quantitatively the impact on crop yields of "weather-up-to-the-present" at successive stages of the crop year. The models were intended to assist commodity specialists in their economic analyses and to serve as one element of a complex, computerized cross-commodity forecasting system under development in the author's working group (Forecast Support Group, Commodity Economics Division, ESCS).

It was apparent from the outset of this research that, because of the large amount of weather data needed for the development of the models, the computer would play a major role in storing and manipulating data. What was not then apparent but became so as the work progressed was that the computerized weather/yield system which was to evolve from this particular project would be sufficiently versatile and general to enable it to serve as a general resource to meet other needs within USDA. Accordingly, a key aim of this article is to acquaint other researchers with the design, operation, and potential of this system, now in operation in USDA.

The four main constituents of the system are: (1) The weather data bank, (2) the economic data bank, (3) the data-manipulation subsystem, and (4) the automated forecasting subsystem.

The Weather Data Bank

Type of Data

Several considerations entered into the early decision as to what type of weather data to employ in the system. Of paramount importance was the fact that the data would have to be transformed into "weather variables," based on both historical and current weather, for use in multivariate regression equations designed to forecast annual crop yields. Because of the need to preserve degrees of freedom, only a few weather variables could be employed in each equation. Thus, weather data representing a large crop-producing region and covering the many months of the growing season would have to be telescoped into a few aggregate figures. This suggested the choice of data which was already spatially and temporally aggregated.

Another consideration was cost. The expense involved in acquiring and storing the historical records of weekly or more-frequent data would have been prohibitive.

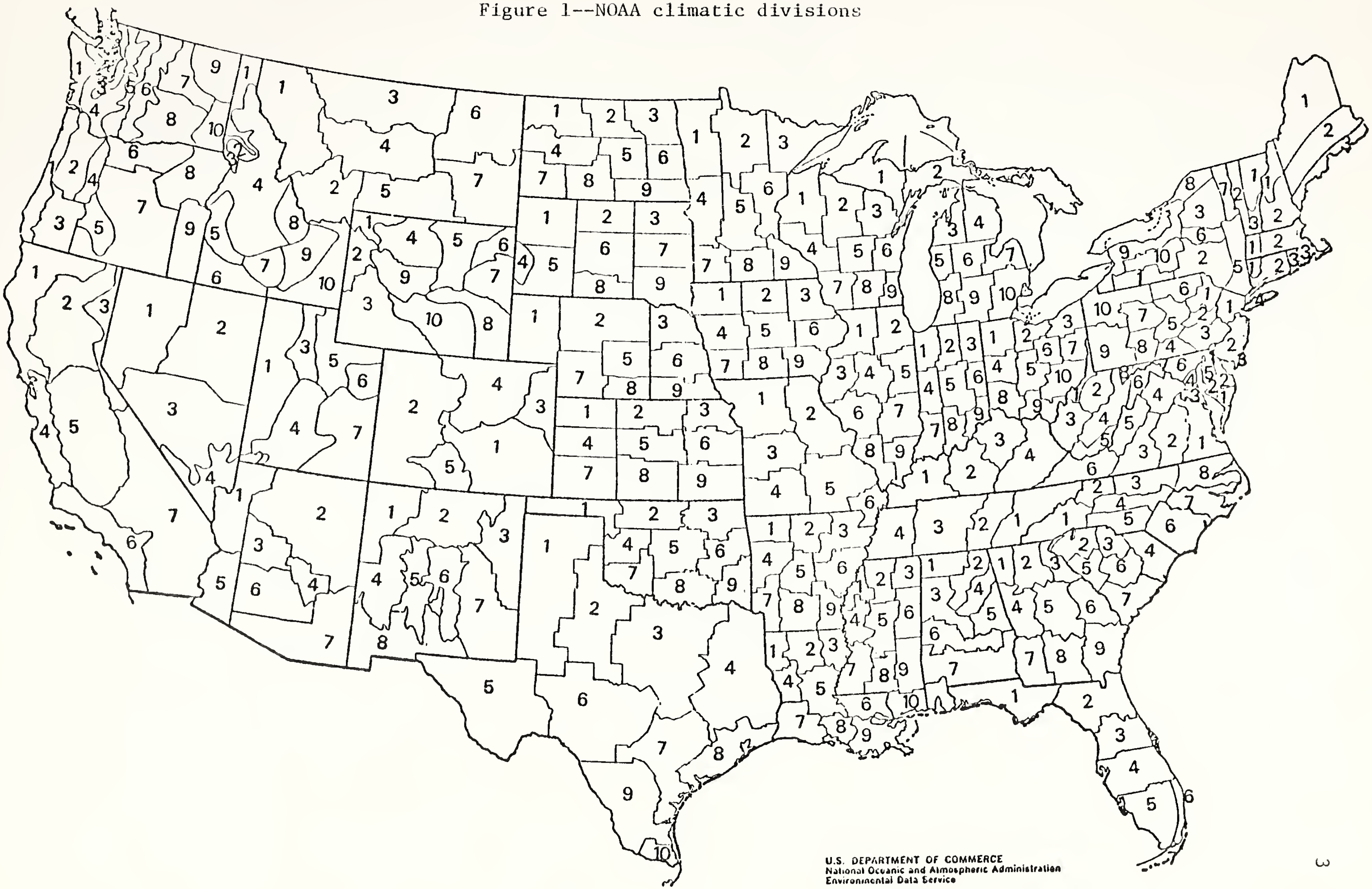
A final consideration was the desire to obtain a collection of historical data which was complete, reasonably "clean," and logically arranged.

A collection of weather records fulfilling these requirements was found to be available from the National Oceanic and Atmospheric Administration (NOAA) and was chosen as the basis of the weather data bank.

The NOAA data is based on a partitioning of each state (and certain other territories) into "climatic divisions"--geographic regions within each of which the climate is relatively homogeneous. For the 48 contiguous states, there are 344 climatic divisions (figure 1); many of these coincide with Crop Reporting Districts as used by ESCS/Statistics. 1/

1/ Formerly the Statistical Reporting Service.

Figure 1--NOAA climatic divisions



The weather data bank consists entirely of NOAA climatic-division data. For each climatic division, the bank currently contains monthly records of (average) temperature and (total) precipitation (January 1931 to present) and the Palmer Drought Index (January 1931 to 1975). In addition, the bank contains 30-year monthly normals, as well as some recent weekly records (used in the weekly updating process, as explained below).

The data are stored on disk in an IBM 370 operated by USDA's Washington Computer Center. They reside in the ESCS/Economics Time-Series Data Management System "TDAM," where they are formatted as monthly or weekly time-series, each series containing the temperature, precipitation, or Palmer Index records for a single climatic division.

The Updating Process

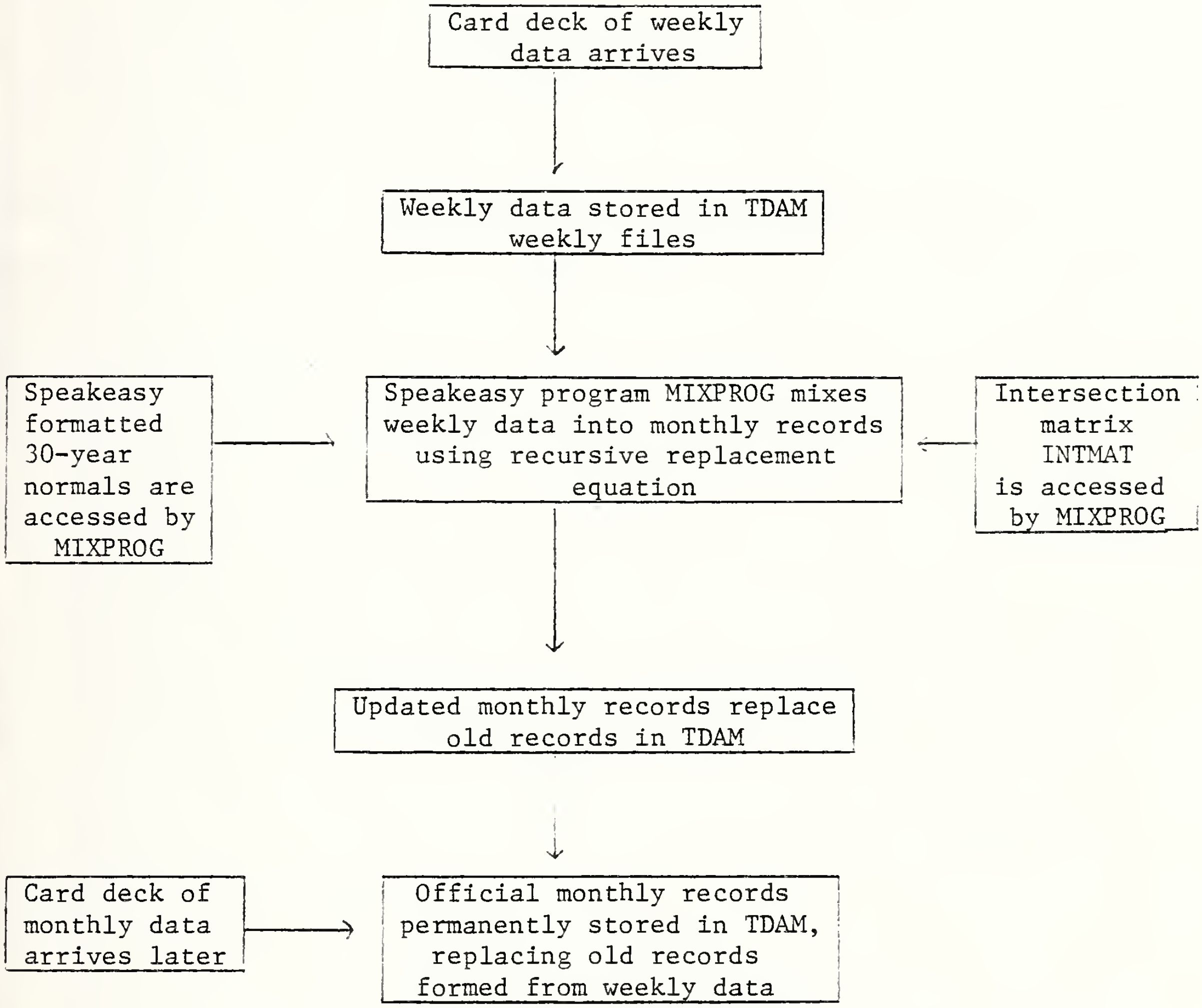
The temperature and precipitation portions of the weather data bank are kept current by a systematic updating process. At regular intervals, NOAA's Federal Climatic Center in Asheville, North Carolina furnishes ESCS with computer card decks containing recent temperature and precipitation records for all climatic divisions. Monthly records are furnished each month throughout the year and cover the month ending six weeks prior. Weekly records are furnished each week during the growing season (approximately March to November) and cover the previous week. Within hours after the receipt of a card deck, its data have been read into the data bank.

As was indicated earlier, the main function of the weekly data is to permit the timely updating of the current month's records. This updating process operates as follows: Within the data bank, each climatic division is assigned a temperature and precipitation "record" for each future month.

The assigned "record" is merely the 30-year normal temperature or precipitation for that month in that climatic division. When, ultimately, data is received covering the first week which intersects a given month, that weekly data is used to revise the monthly record, in effect replacing between-one-and-seven days of normal weather with actual weather. When the next week's data is received, more days of normal weather in the monthly record are replaced by actual weather. This process of weekly replacement continues until the monthly record consists entirely of an aggregation of actual (weekly) data. NOAA's weekly weather data, however, is intended only to be provisional and is subject to revision. Thus, six weeks later, when NOAA's "official" record for the given month is received, that record permanently replaces the "aggregated" record just described (figure 2).

The logical requirements of an automated, weekly, recursive replacement process such as the one just described are more complex than at first might be assumed. On the one hand, one desires that the controlling software be compact, logically simple, and unrequiring of internal change once installed. On the other hand, the replacement process must account for the fact that different months have different lengths, and leap years affect the length of February; that there is no prior certainty as to in which week in a crop year NOAA will begin publication of its weekly data; that some weeks intersect (and thus must trigger revision of) two months; importantly, that the calendar, and thus the entire interweaving of weeks and months, changes yearly; and, finally, that weekly temperatures and weekly precipitations must be aggregated into monthly figures in different manners, since temperature is in general expressed as an average, precipitation as a sum.

Figure 2--Weather data bank updating process



Despite this labyrinth of detail, essentially a single, simple equation was developed which, applied recursively to each climatic division, accomplishes the weekly replacement process (figure 3). In this approach, the underlying algorithm--essentially an abstract "model" of the replacement process--remains unchanged through different weeks, months, and years. The only notable requirement placed on the user is that, at the beginning of each crop year, he submit the simple "intersection matrix" INTMAT whose k, j element is the number of days common to month k and NOAA week j. (Actually, INTMAT can be generated automatically, once the user has designated which week is NOAA's "Week 1" for that crop year.) In effect, this approach permits all calendar-related detail to be confined within INTMAT, exterior to the permanent software.

Since applied economic research not infrequently involves the consolidation of incommensurably-timed data-series, it might be mentioned that the approach described here--of capturing the essence of the mixing process in a fixed, abstract algorithm while confining all ad hoc time-relationships within a simple, external "intersection matrix"--should be applicable in other contexts.

The Economic Data Bank

Though weather is an important determinant of crop yields, it is by no means the only influence. Economic factors, in particular, play a measurable role. If fertilizer prices increase and farmers thus apply less fertilizer to their crops, yield may decrease. Or, if crop prices decrease and as a result farmers plant fewer acres, they may well remove from production the least-productive land first, so that yield (which, after all, is by definition the ratio of quantity-of-crop-produced to land-area used) may increase.

Figure 3--Recursive data-replacement equation

$$\text{NEWVAL} = \text{OLDVAL} + \text{INTMAT}(k,j) \cdot \left[\frac{\text{WEEKVAL}}{\text{TIMETYP}} - \frac{\text{MONNORM}}{\text{MONLENGTH}}\right]$$

This equation is used for each climatic division to mix temperature or precipitation data for week j into the provisional record for month k.

Variables:

NEWVAL is monthly record after mixing.
OLDVAL is monthly record before mixing.
INTMAT(k,j) is number of days common to month k and NOAA week j.
WEEKVAL is record for week j.
MONLENGTH is number of days in month k.
TIMETYP is 7 for precipitation, MONLENGTH for temperature.
MONNORM is 30-year monthly normal for climatic division (temperature or precipitation, as the case may be).

Consistent with the responsibilities of ESCS/Economics for economic research and analysis, it was anticipated from the beginning of the project that the regression equations ultimately developed would reflect the impact on yield not only of weather, but also of economic factors. However, as preliminary equations were developed, often in an atmosphere of urgency, the lack of a complete, accurate, and readily available set of state-level economic data proved a constant irritant. A good deal of data had to be copied hurriedly by hand from diverse statistical compendia and then typed into computer storage. The continuation of this situation was unacceptable, and it was determined that a definitive, computerized bank of state-level economic data would be necessary. Through several weeks of intense effort, by combining, correcting, and extending data-series culled from various fragmentary sources, this bank was created and made an integral part of the developing system.

At its current stage of development, the economic data bank consists of annual records (1939 to present) of acres planted, acres harvested, and yield per harvested acre, for the U.S. and the 48 contiguous states. Corn, soybeans, sorghum, barley, oats, and several categories of wheat are represented. It is anticipated that further classes of data will be introduced as time permits.

Aside from their use in the creation of purely economic variables, the acreages data also play an indispensable role in the creation of weather variables, for temperature and precipitation figures for a region are typically "weighted" by planted- or harvested acreages in order that weather effects be related fairly to their agricultural consequences.

The economic data, unlike the weather data, reside in (disk) files which are directly manipulated through the computer language "Speakeasy" (about which, more later). They are stored as matrix-like "boxes," one for each family of data, with rows corresponding to years and columns corresponding to states or U.S. Hence, for example, CRNBXAP ("corn box--acres planted") is an array of columns and rows containing all of the annual state- and U.S. data on acres planted for corn.

The economic data bank is specifically designed to exploit the powerful capabilities of Speakeasy. Each data-box in its entirety, or portions of it, may be accessed and manipulated with unusual ease.

Figure 4 illustrates the operation of the data bank, in an interactive ("on-line") mode, by a user who wishes to obtain the historical annual ratios of harvested- to planted acreages for corn, for the combined region formed by Illinois, Iowa, Indiana, Ohio, and Nebraska. First, the corn acreage data is brought from storage into the "workspace." Then, in a single swoop, the states are selected, the regional totals of harvested and planted acreages formed, and the annual ratios computed. Finally, a printout of the results is generated. Observe that, within the data-boxes, states are identified by simple Postal Service abbreviations.

An important feature of the data bank is the ease with which data may be updated, a process currently implemented manually. Though there are hundreds of time-series to be updated several times yearly, perhaps only two man-days per year are required for data maintenance. Essentially, to update a data-box, the operator simply types the new figures in the same order as they appear in the publication being used; the system already "knows" which states correspond. Then, a brief command serves to update all time-series in the data-box simultaneously.

Figure 4--Example of economic data bank use*

```
:_GET CRNBXAH; GET CRNBXAP
:_RATIO=SUMROWS(CRNBXAH(,(IL IA IN OH NE)))/SUMROWS(CRNBXAP(,(IL IA IN OH NE)))
:_TABULATE(YEAR RATIO)

 YEAR  RATIO      YEAR  RATIO      YEAR  RATIO      YEAR  RATIO
.....  .....     .....  .....     .....  .....     .....  .....
 1939   .90       1949   .96       1959   .96       1969   .94
 1940   .89       1950   .95       1960   .96       1970   .94
 1941   .95       1951   .94       1961   .96       1971   .95
 1942   .95       1952   .96       1962   .95       1972   .94
 1943   .93       1953   .95       1963   .95       1973   .95
 1944   .94       1954   .94       1964   .95       1974   .92
 1945   .93       1955   .91       1965   .95       1975   .94
 1946   .95       1956   .88       1966   .95       1976   .94
 1947   .92       1957   .95       1967   .94       1977   .94
 1948   .96       1958   .95       1968   .94       1978   .94

:_
```

* In this and subsequent figures, user input is framed to distinguish it from computer output.

Data-Manipulation Subsystem

Early experience in using the weather data bank demonstrated that a special data-access-and-manipulation system would have to be developed if the user were not to be overwhelmed by the sheer mass of data and number of operations required in ordinary applications.

The derivation of a typical corn-yield regression equation offers a case in point. A regional approach to corn-yield estimation might easily involve six states, fifty climatic divisions, and (since temperature and precipitation are both considered) one hundred weather time-series. The series would have to be weighted by published NOAA "area coefficients" to compensate for differences in area between climatic divisions of a state, and by further coefficients to account for differences in area between states. Once the series had been combined to form two regional aggregate series (one for temperature, one for precipitation), perhaps half-a-dozen different monthly subseries would then have to be extracted from each for statistical testing in the desired regression.

In the initial phase of the project, the need for immediate, usable results made procedures of this explicit complexity unavoidable. However, as time permitted, a sequence of increasingly powerful data-manipulation routines was developed, culminating in the data-manipulation subsystem now in place.

The common language of this subsystem, and the key to its flexibility and sophistication, is the computer language "Speakeasy" mentioned earlier. Developed in the 1960's at the Argonne National Laboratories, Speakeasy was originally intended for atomic-energy research. However, its impressive capabilities in performing high-level mathematical, logical, and editorial operations, especially on large arrays of numbers or characters, prompted

its adoption in other disciplines. The version maintained in ESCS ("Fedeasy") contains econometric routines added by the Federal Reserve Bank.

The heart of the data-manipulation subsystem is the Speakeasy program UTREV ("utility-revised"). UTREV is capable of manipulating weather records for any chosen groupings of years, months, states, and state climatic divisions. Its usefulness is most easily grasped through several actual examples of its operation as experienced interactively by the user at the computer terminal:

Example 1

As part of a study on corn yields, the user wishes to create annual time-series representing early-season (September-May) and late-season (June-August) precipitation for the Corn Belt.

He enters (figure 5) the weather type ("PRCP"); the state names ("IA," "NE," etc.); for each state, the climatic divisions where corn is produced ("ALL;" "3 6 9;" etc.); the full list of months considered ("SEP," "OCT," etc.); and the years ("1931," "1977"). He then iteratively selects two groupings of months for aggregation and assigns names to the resulting data arrays.

At this stage, the user has created separate annual early- and late-season series for each state. With a few brief Speakeasy commands (currently being incorporated into a single command in UTREV), he can now combine these series into regional series, each of which represents the total early- or late-season precipitation for the entire Corn Belt (figure 6). Alternatively, with another command, he can weight the separate state

Figure 5--Using UTREV for Corn Belt precipitation

```
SPECIFY WEATHER TYPE (TEMP, PRCP, OR PDI)
WTYPE = PRCP
ENTER STATE NAME ABBREVIATIONS
STATNAMS = IA NE MN IL IN OH WI MO SD MI
ENTER "ALL" OR LIST CHOSEN CLIMATIC SUBDIVISIONS IN
ORDER WITHOUT LEADING ZEROES
SUBDVSIA = ALL
SUBDVSNE = 3 6 9
SUBDVSMN = 4 5 6 7 8 9
SUBDVSIL = ALL
SUBDVSIN = ALL
SUBDVSOH = ALL
SUBDVSWI = 4 5 6 7 8 9
SUBDVSMO = 1 2 3 4 6
SUBDVSSD = 2 3 6 7 8 9
SUBDVSMI = 5 6 7 8 9 10
ENTER MONTHS IN ORDER
MONVALS = SEP OCT NOV DEC JAN FEB MAR APR MAY JUN JUL AUG
ENTER STARTING AND ENDING YEARS FOR FIRST MONTH
STARTYR = 1931
LASTYR = 1977
WANT TO COMBINE MONTHS?
ANS = YES
ENTER MONTHS TO BE COMBINED IN ORDER
CHOICES = SEP OCT NOV DEC JAN FEB MAR APR MAY
WITHOUT BLANKS, ENTER NAME OF FORTHCOMING ACCUMULATION MATRIX EARLY
EARLY IS A 47 BY 10 REAL ARRAY
WANT TO COMBINE MONTHS AGAIN?
ANS = YES
ENTER MONTHS TO BE COMBINED IN ORDER
CHOICES = JUN JUL AUG
WITHOUT BLANKS, ENTER NAME OF FORTHCOMING ACCUMULATION MATRIX LATE
LATE IS A 47 BY 10 REAL ARRAY
WANT TO COMBINE MONTHS AGAIN?
ANS = NO
:_
```

Figure 6--Corn Belt precipitation series

CROPYEAR	EARLYREG	LATEREG
....	....	....
1932	23.88	11.70
1933	23.14	7.24
1934	13.69	8.80
1935	24.84	11.85
1936	17.18	5.90
1937	24.86	10.61
1938	22.62	11.37
1939	20.41	11.93
1940	15.62	11.22
1941	17.03	9.99
1942	26.77	12.53
1943	23.43	12.00
1944	21.10	11.91
1945	22.59	11.49
1946	22.24	10.40
1947	24.03	11.47
1948	20.93	11.03
1949	21.56	11.35
1950	24.53	11.81
1951	22.12	14.03
1952	22.98	11.55
1953	18.00	9.78
1954	16.55	11.40
1955	20.34	9.87
1956	18.31	11.45
1957	19.61	12.46
1958	18.91	13.88
1959	20.70	9.73
1960	23.87	11.29
1961	19.77	10.74
1962	24.34	11.36
1963	17.67	10.40
1964	18.08	10.51
1965	20.98	11.48
1966	22.38	9.95
1967	19.82	11.03
1968	21.48	11.58
1969	23.27	12.55
1970	21.30	9.83
1971	22.23	9.42
1972	22.67	11.41
1973	29.07	10.45
1974	26.88	10.21
1975	22.21	11.84
1976	20.56	7.64
1977	16.94	12.92
1978	24.23	11.91

: _

series with state harvested- or planted acreages for corn, arriving at series which represent the total early- or late-season precipitation for the Corn Belt farmland actually utilized for corn production (figure 7).

Example 2

In order to investigate the relationship between low temperatures and cattle death losses, the user wishes to obtain annual time-series for March and April average temperatures over a major livestock-producing region. He enters the weather type, the states, the respective climatic divisions, and the months and years (figure 8) and calls for separate March and April (state) series to be formed (figure 9 displays these separate series for March). These series are then aggregated to form the regional series desired (figure 10). Alternatively, the state series could be weighted by state herd sizes, if desired.

The next example highlights one of the most useful features of the data-manipulation subsystem--the "probability operator."

Example 3

Concerned about the prospects for the sorghum crop in the Texas Panhandle, the user wishes to determine the probabilities (based on historical frequencies) of various weather events for the month of March (such as the probability of a "hot and dry" March).

As before, he enters the appropriate information (first for precipitation, then for temperature), choosing Texas climatic divisions 1 and 2 to represent the Panhandle (figure 11). He thereby obtains annual time-series for March precipitation and temperature for the Panhandle (figure 12).

Figure 7--Acreage-weighted Corn Belt precipitation series

CROPYEAR	EARLYWT	LATEWT	CROPYEAR	EARLYWT	LATEWT
....	....	......	....	....	......
1939	20.06	12.14	1959	21.01	9.89
1940	15.08	10.94	1960	23.79	11.71
1941	16.94	10.11	1961	19.23	10.90
1942	26.67	12.67	1962	24.71	11.91
1943	22.18	12.38	1963	17.38	10.92
1944	21.51	12.56	1964	18.08	10.85
1945	22.22	11.60	1965	21.59	11.74
1946	21.69	10.78	1966	22.67	9.89
1947	23.87	11.70	1967	19.81	11.53
1948	20.66	11.27	1968	21.39	11.56
1949	21.79	11.50	1969	23.76	13.18
1950	24.26	12.11	1970	21.44	10.25
1951	22.36	14.83	1971	22.30	9.55
1952	22.70	12.02	1972	22.67	11.97
1953	17.72	9.73	1973	29.09	10.73
1954	16.65	11.83	1974	26.97	10.39
1955	19.46	9.69	1975	21.89	11.89
1956	17.54	11.34	1976	20.63	7.66
1957	19.36	12.99	1977	16.87	13.50
1958	18.86	14.50	1978	23.13	10.23

:_

Figure 8--Using UTREV in cattle death-loss study

```
SPECIFY WEATHER TYPE (TEMP, PRCP, OR PDI)
WTYPE = TEMP
ENTER STATE NAME ABBREVIATIONS
STATNAMS = WA MT WY CO ND SD NE KS
ENTER "ALL" OR LIST CHOSEN CLIMATIC SUBDIVISIONS IN
ORDER WITHOUT LEADING ZEROES
SUBDVSWA = 7 8 9 10
SUBDVSMT = 3 4 5 6 7
SUBDVSWY = 2 4 5 6 7 8
SUBDVSCO = 1 3 4
SUBDVSND = 1 4 7 8
SUBDVSSD = 1 4 5
SUBDVSNE = 1 2 7
SUBDVSKS = 1 4 7
ENTER MONTHS IN ORDER
MONVALS = MAR APR
ENTER STARTING AND ENDING YEARS FOR FIRST MONTH
STARTYR = 1931
LASTYR = 1978
WANT TO COMBINE MONTHS?
ANS = YES
ENTER MONTHS TO BE COMBINED IN ORDER
CHOICES = MAR
WITHOUT BLANKS, ENTER NAME OF FORTHCOMING ACCUMULATION MATRIX MARTEMP
MARTEMP IS A 48 BY 8 REAL ARRAY
WANT TO COMBINE MONTHS AGAIN?
ANS = YES
ENTER MONTHS TO BE COMBINED IN ORDER
CHOICES = APR
WITHOUT BLANKS, ENTER NAME OF FORTHCOMING ACCUMULATION MATRIX APRTEMP
APRTEMP IS A 48 BY 8 REAL ARRAY
WANT TO COMBINE MONTHS AGAIN?
ANS = NO
:_
```

Figure 9--March temperature series by state

MARTEMP

YEAR	WA	MT	WY	CO	ND	SD	NE	KS
1931	41.9	31.6	31.0	33.4	26.4	31.0	33.6	37.5
1932	39.8	22.8	25.0	31.0	19.8	25.1	29.0	36.4
1933	40.0	34.7	34.2	39.8	30.3	35.3	38.3	45.7
1934	46.8	33.5	36.5	41.3	27.8	33.4	38.1	44.1
1935	38.6	27.9	32.5	40.8	27.1	35.5	40.6	49.1
1936	39.6	30.4	31.3	39.2	26.8	34.3	38.3	46.3
1937	42.6	29.9	30.5	34.8	26.9	30.7	34.0	39.5
1938	40.9	33.3	35.2	40.4	32.7	38.0	41.7	49.1
1939	41.8	31.8	33.6	37.4	24.9	34.5	38.9	43.5
1940	45.2	33.6	36.5	39.6	26.1	32.6	37.8	44.3
1941	46.3	32.7	32.7	35.2	26.5	32.2	34.3	40.2
1942	41.2	32.5	31.0	34.0	29.7	33.8	36.0	42.9
1943	37.5	22.9	25.4	34.2	17.6	25.1	30.6	39.3
1944	39.0	24.0	27.4	33.3	16.7	22.6	29.6	37.7
1945	39.3	34.3	33.7	39.9	33.4	36.5	41.4	48.2
1946	41.8	39.2	38.7	41.9	36.5	41.9	44.2	49.2
1947	44.1	25.2	30.2	34.5	22.6	26.7	32.8	38.6
1948	37.8	25.9	27.8	29.2	19.1	28.5	31.2	34.3
1949	40.3	28.3	32.1	37.6	23.1	29.9	35.3	41.7
1950	37.9	24.9	29.5	36.6	19.6	24.7	31.8	40.4
1951	35.5	18.3	27.2	34.4	11.4	22.2	30.4	38.8
1952	39.1	23.2	27.6	32.5	17.1	23.2	30.0	37.0
1953	42.0	33.6	36.0	41.8	26.8	34.4	40.2	46.9
1954	37.6	24.4	27.4	34.7	22.7	26.3	31.5	39.7
1955	33.6	23.2	25.7	36.3	21.3	28.5	33.9	41.3
1956	37.8	31.1	32.0	37.7	24.7	33.6	37.0	43.6
1957	40.0	31.4	33.5	37.0	27.5	32.2	34.8	40.9
1958	40.4	27.9	29.1	30.0	24.6	29.0	26.7	29.6
1959	41.0	35.2	33.6	36.4	32.5	36.4	37.0	42.1
1960	40.0	28.2	32.9	35.3	19.2	26.2	29.0	35.0
1961	41.9	37.1	36.0	37.4	33.7	38.0	38.1	42.8
1962	37.7	25.2	29.7	33.7	21.9	27.2	31.0	39.6
1963	42.3	37.3	35.2	37.7	34.5	39.3	39.9	46.4
1964	38.0	26.9	26.7	31.8	21.8	28.8	31.2	39.2
1965	37.6	17.7	21.1	27.0	13.6	20.0	24.5	32.5
1966	40.9	34.7	35.7	40.3	31.9	35.4	39.0	45.6
1967	40.1	27.5	35.3	42.1	26.6	34.8	40.7	47.2
1968	44.3	38.9	37.3	38.1	35.1	39.1	41.2	46.0
1969	39.7	25.1	28.0	29.5	18.9	26.3	28.6	32.3
1970	39.9	27.1	28.5	31.7	19.6	27.2	30.8	36.2
1971	36.3	30.1	30.8	36.4	26.5	31.3	34.1	42.2
1972	42.6	35.2	38.8	42.5	25.8	36.1	41.2	48.9
1973	42.7	37.1	33.5	35.3	36.2	37.9	39.5	43.6
1974	40.7	32.1	35.4	41.1	25.6	36.2	39.9	46.4
1975	37.3	26.7	30.1	34.8	21.2	27.1	31.4	39.0
1976	37.2	30.1	30.2	36.3	25.3	32.6	35.9	43.0
1977	40.3	33.8	31.7	35.5	33.4	35.9	36.2	44.5
1978	43.2	31.5	34.7	39.1	26.9	33.3	35.8	42.9

:_

Figure 10--March-April temperature series for livestock-producing region

YEAR	TMARREG	TAPRREG	YEAR	TMARREG	TAPRREG	YEAR	TMARREG	TAPRREG
1931	32.8	45.9	1947	30.6	44.5	1963	38.3	45.8
1932	27.4	47.6	1948	28.4	47.4	1964	29.6	45.0
1933	36.7	43.2	1949	32.6	49.1	1965	22.9	46.4
1934	37.0	49.0	1950	29.8	41.6	1966	37.3	42.0
1935	34.8	41.5	1951	25.9	42.0	1967	35.2	43.7
1936	34.6	43.9	1952	27.8	48.4	1968	39.5	42.9
1937	32.8	45.0	1953	37.1	40.5	1969	27.9	49.1
1938	37.6	45.9	1954	29.4	45.5	1970	29.5	40.5
1939	35.0	46.9	1955	29.1	46.7	1971	32.8	45.1
1940	36.4	43.6	1956	34.0	42.2	1972	38.3	44.7
1941	34.4	45.8	1957	34.1	42.4	1973	37.5	41.9
1942	34.3	48.6	1958	29.2	43.9	1974	36.2	46.5
1943	27.9	50.5	1959	36.2	44.3	1975	30.3	39.7
1944	28.0	43.8	1960	30.6	46.3	1976	33.0	47.0
1945	37.4	41.0	1961	37.7	42.2	1977	35.5	49.3
1946	41.0	51.7	1962	29.7	48.2	1978	35.1	46.2

Figure 11--Using UTREV in Texas sorghum study

```
SPECIFY WEATHER TYPE (TEMP, PRCP, OR PDI)
WTYPE = PRCP
ENTER STATE NAME ABBREVIATIONS
STATNAMS = TX
ENTER "ALL" OR LIST CHOSEN CLIMATIC SUBDIVISIONS IN
ORDER WITHOUT LEADING ZEROES
SUBDVSTX = 1 2
ENTER MONTHS IN ORDER
MONVALS = MAR
ENTER STARTING AND ENDING YEARS FOR FIRST MONTH
STARTYR = 1931
LASTYR = 1977
WANT TO COMBINE MONTHS?
ANS = NO
:_
```

```
SPECIFY WEATHER TYPE (TEMP, PRCP, OR PDI)
WTYPE = TEMP
ENTER STATE NAME ABBREVIATIONS
STATNAMS = TX
ENTER "ALL" OR LIST CHOSEN CLIMATIC SUBDIVISIONS IN
ORDER WITHOUT LEADING ZEROES
SUBDVSTX = 1 2
ENTER MONTHS IN ORDER
MONVALS = MAR
ENTER STARTING AND ENDING YEARS FOR FIRST MONTH
STARTYR = 1931
LASTYR = 1977
WANT TO COMBINE MONTHS?
ANS = NO
:_
```

Figure 12--Texas Panhandle precipitation and temperature series

YEAR	PANPRCP	PANTEMP	YEAR	PANPRCP	PANTEMP	YEAR	PANPRCP	PANTEMP
1931	1.25	46.1	1947	1.17	45.8	1963	.45	53.9
1932	.29	46.2	1948	.73	46.3	1964	.49	49.5
1933	.51	52.9	1949	.84	50.8	1965	.45	42.3
1934	1.85	50.1	1950	.05	50.6	1966	.34	53.4
1935	1.00	56.0	1951	.80	49.8	1967	.75	57.1
1936	.45	54.6	1952	.59	49.1	1968	1.61	50.1
1937	1.45	46.3	1953	1.18	56.2	1969	1.70	41.8
1938	1.33	56.3	1954	.12	49.3	1970	2.19	45.4
1939	.84	53.6	1955	.41	52.0	1971	.06	50.5
1940	.14	52.7	1956	.06	52.9	1972	.21	56.1
1941	2.11	46.5	1957	1.70	50.0	1973	3.48	51.2
1942	.85	50.0	1958	2.10	42.3	1974	1.12	57.0
1943	.65	47.8	1959	.34	50.3	1975	.44	48.3
1944	.46	48.9	1960	.61	45.5	1976	.61	51.6
1945	.90	54.3	1961	2.08	51.5	1977	.83	51.3
1946	.73	54.2	1962	.49	48.6			

:_

To obtain the desired probabilities, he defines "HOT" as the event that the March temperature over the Panhandle is greater than ("GT") average and defines "DRY" analogously (figure 13). (Other definitions would of course be possible.) He now merely enters his probability questions in a natural form as shown, and the system responds with the answers (in this case, concluding with the probability of obtaining more than two inches of precipitation).

In addition to UTREV, the data-manipulation subsystem contains a variety of other specialized programs and auxiliary data sets. Notable among the latter are collections of state areas, climatic-division area coefficients, and 30-year monthly weather normals. Furthermore, operation of the subsystem implies immediate access to the full range of standard Speakeasy operations, permitting extensive statistical and mathematical analyses to be conducted at any stage of a study.

The Automated Forecasting Subsystem

The automated yield-forecasting subsystem represents the culmination of all the development efforts previously described. It consists of Speakeasy programs which generate yield forecasts for major commodities, using regression equations reflecting both economic factors and "weather-up-to-the-present." Using modified non-interactive versions of UTREV, these programs automatically select commodity-related weather records from the weather data bank, weight them with acreage data from the economic data bank to form aggregate weather variables, and apply previously determined regression coefficients to these and selected macroeconomic variables (stored elsewhere) to produce crop-yield forecasts. Figure 14 schematizes

Figure 13--Application of the probability operator

```
: HOT=PANTEMP.GT.MEAN(PANTEMP)
:
: DRY=PANPRCP.LT.MEAN(PANPRCP)
:
: PROB(HOT.AND.DRY)
PROB(HOT.AND.DRY)  =   .36
:
: PROB(HOT.OR.DRY)
PROB(HOT.OR.DRY)  =   .79
:
: PROB(HOT)
PROB(HOT)  =   .49
:
: PROB(DRY)
PROB(DRY)  =   .66
:
: PROB(PANPRCP.GT.2)
PROB(PANPRCP.GT.2)  =   .11
:
```

the operation of the automated forecasting subsystem in its relationship to the entire weather/yield system.

Figure 15 outlines a "first-generation" soybean-yield equation whose forecasts have been automated. Early in the project, the development of this and similar regression equations combining aggregate weather data with macroeconomic factors served to confirm the usefulness of further effort. Figure 16 shows how the forecast is obtained interactively.

Epilogue

The weather/yield system has developed into an information, research, and forecasting tool capable of responding to a variety of needs with only minimal requirements of the user. The weather data bank itself has already performed an auxiliary role as a basic data source for a major USDA study (7).

Other applications and extensions can be envisioned, ranging from improved estimation of pig death losses in econometric models to better understanding of the economic impacts of weather on rural populations. Of particular significance could be the broadening of the system to include foreign countries. This would provide a foundation for strengthening yield-modelling efforts in the context of the world agricultural economy.

Figure 14--Operation of automated forecasting subsystem

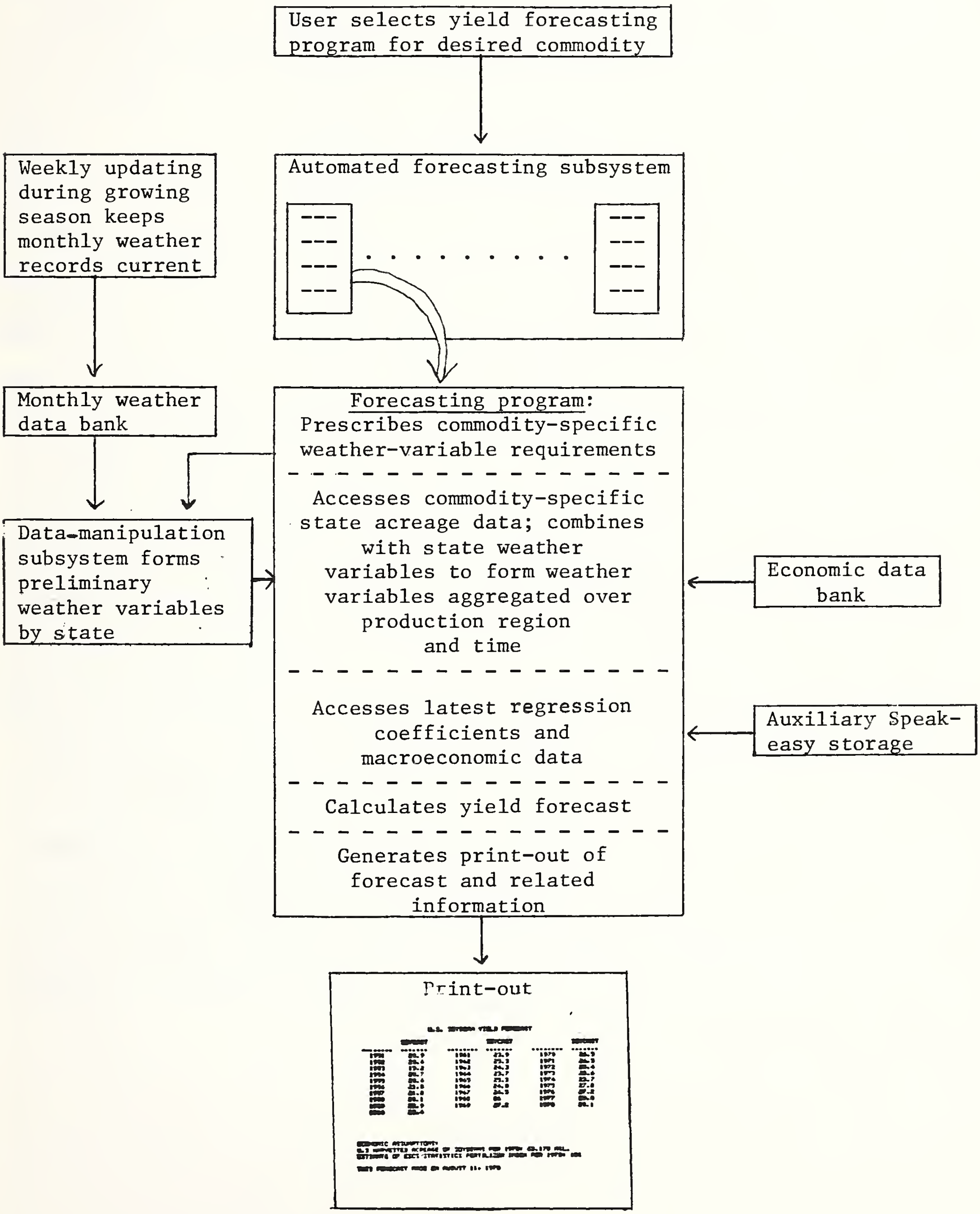


Figure 15--U.S. soybean-yield regression results

Variable	Estimated coefficient	Standard error	T statistic
b_o (constant)	13.20	1.98	6.68
SOYSH	-0.04	0.09	-0.42
FERTM	-0.01	0.01	-1.02
INDEXSM	2.90	1.73	1.67
INDEXJA	4.08	1.01	4.04
DUM74	-4.75	1.07	-4.43
TREND	0.41	0.15	2.79

R squared = 0.92

R squared (corrected) = 0.90

Durbin-Watson statistic = 2.11

Standard error = 0.89

Data from years 1951-76

Regression Variables:

SOYSH is the U.S. harvested acreage of soybeans in millions of acres.

FERTM is a fertilizer price index.

INDEXSM is an early-season precipitation index.

INDEXJA is a late-season precipitation index.

DUM74 is a dummy variable equaling 1 in 1974 and 0 in other years.

TREND is a linear time trend.

Figure 16--Use of the automated forecasting subsystem: generating a soybean-yield forecast

```
TSO SPEAKEASY 3 MU+ 11:15 AM AUGUST 11, 1978
:_SIZE=175
:_GET SOYPROG
:_SOYPROG
EXECUTION STARTED
```

****

U.S. SOYBEAN YIELD FORECAST

	SOYCAST		SOYCAST		SOYCAST
1951	20.9	1961	23.9	1970	26.5
1952	20.4	1962	25.3	1971	26.5
1953	19.2	1963	24.3	1972	28.4
1954	20.7	1964	23.7	1973	28.6
1955	20.4	1965	25.3	1974	23.7
1956	21.8	1966	24.8	1975	27.8
1957	21.8	1967	24.5	1976	27.2
1958	24.1	1968	26	1977	29.8
1959	22.9	1969	27.2	1978	29.1
1960	23.4				

```
ECONOMIC ASSUMPTIONS:
U.S HARVESTED ACREAGE OF SOYBEANS FOR 1978: 63.178 MIL.
ESTIMATE OF ESCS/STATISTICS FERTILIZER INDEX FOR 1978: 181

THIS FORECAST MADE ON AUGUST 11, 1978
```

****

```
MANUAL MODE
:_
```

REFERENCES

1. Climatological Data (previously published under the title Climatic Data for the United States by Sections). U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Data Service, all issues, 1931-1979.

2. Cohen, Stanley and Steven C. Pieper, The SPEAKEASY-3 REFERENCE MANUAL, LEVEL MU. Argonne National Laboratory report ANL-8000, rev. 2, 1977.

3. McQuigg, James D., Economic Impacts of Weather Variability (report prepared for National Science Foundation). Dept. of Atmospheric Science, Univ. of Missouri-Columbia, August 1975.

4. Nix, James E., "Severe Winter Weather Affects Livestock Industry." Weekly Weather and Crop Bulletin, U.S. Dept. of Commerce and U.S. Dept. of Agriculture, Vol. 66, No. 11, March 13, 1979.

5. Thompson, Louis M., "Weather and Technology in the Production of Corn in the U.S. Corn Belt." Agronomy Journal, Vol. 61, pp. 453-456, 1969.

6. ____________, "Weather and Technology in the Production of Soybeans in the Central United States." Agronomy Journal, Vol. 62, pp. 232-236, 1970.

7. Weather and Water Allocation Study (report prepared for Congress and the President pursuant to the Food and Agriculture Act of 1977). U.S. Dept. of Agriculture, 1979.

8. Weiss, Michael D., "Weather and Crop Yield Analysis System." Proceedings of the 6th Annual Speakeasy Conference (Chicago, Illinois, August 17-18, 1978), Argonne National Laboratory, 1978.